Cognitive Chain Optimization: A Dual-Process Reinforcement Framework for Explainable LLM Reasoning

Henrik Nieminen
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
henriknieminen@binghamton.edu

Logan Day
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
hellologan@missouri.edu

Ajay R. Dutta
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
ajay1996@ku.edu

Abstract

The rapid advancement of large language models (LLMs) has produced systems capable of sophisticated language generation, yet their reasoning processes remain largely opaque, presenting significant challenges for trust, verification, and governance in high-stakes applications. This paper proposes a novel framework, Cognitive Chain Optimization (CCO), which integrates dual-process cognitive theory with reinforcement learning to construct explainable reasoning trajectories in LLMs. The framework distinguishes between two complementary reasoning modalities: an intuitive, associative System One for rapid heuristic generation, and a deliberative, analytical System Two for structured logical verification. Reinforcement learning is employed not merely to optimize final outputs, but to shape the cognitive chain itself, rewarding intermediate reasoning steps that are both accurate and interpretable. We argue that this approach addresses fundamental structural trade-offs between reasoning depth, computational efficiency, and model transparency. The paper examines the architectural implications of implementing such a dual-system within transformer-based models, considering both modular and integrated design patterns. We analyze the governance and policy dimensions of deployable explainable AI systems, emphasizing the need for auditability, fairness, and robustness in reasoning chains. Sustainability challenges related to the computational overhead of dual-process inference are discussed alongside potential infrastructure solutions, including hierarchical caching and context-aware resource allocation. Cross-domain comparisons with human cognitive science and classical symbolic AI illuminate the theoretical foundations of the proposed framework. A case illustration in medical diagnostic reasoning demonstrates the practical viability of CCO. Finally, we outline forward-looking perspectives on the evolution of explainable reasoning agents and their role in socio-technical infrastructures. This paper contributes a systems-level perspective on aligning LLM reasoning processes with human interpretability requirements, offering a foundation for future research in explainable artificial intelligence, cognitive architectures, and AI governance.

Keywords

explainable AI, large language models, dual-process theory, reinforcement learning, cognitive architecture, reasoning transparency, AI governance, socio-technical systems.

1. Introduction

The emergence of large language models as general-purpose reasoning engines has transformed the landscape of artificial intelligence, yet a fundamental tension persists between their remarkable capabilities and the opacity of their internal operations. Modern LLMs, built upon transformer architectures trained on vast corpora, generate coherent and often insightful responses without providing clear access to the reasoning pathways that produce them [1], [2]. This opacity raises critical concerns for deploying these systems in domains where decisions carry significant consequences, including healthcare, legal reasoning, financial analysis, and public policy. The challenge is not merely technical but deeply socio-technical, intersecting with questions of accountability, trust, and institutional governance [3], [4].

Explainable AI has emerged as a necessary counterweight to the black-box nature of deep learning systems. However, many existing approaches to explainability operate post-hoc, generating rationalizations after a decision has been made rather than revealing the actual computational processes that led to that decision [5], [6]. Such methods, while practically useful, risk producing explanations that are plausible but not faithful to the underlying model dynamics. This paper argues for a fundamentally different approach: designing reasoning architectures that are intrinsically explainable, where the reasoning process is both accessible and auditable by design. We propose Cognitive Chain Optimization, a framework that marries dual-process cognitive theory with reinforcement learning to produce structured, transparent reasoning chains in LLMs.

The theoretical foundation for this work draws from Kahneman's dual-process model of human cognition, which distinguishes between fast, automatic, intuitive thinking (System One) and slow, deliberate, analytical thinking (System Two) [7]. In human reasoning, these two systems interact continuously, with System One generating rapid judgments that System Two may subsequently monitor, correct, or override. We argue that LLMs, despite their fundamentally different computational substrate, can benefit from a similar architectural separation of reasoning modalities. By implementing distinct pathways for heuristic generation and logical verification within a single model, we create a cognitive chain that is both efficient and accountable.

Reinforcement learning serves as the mechanism for optimizing this cognitive chain. Unlike standard RL approaches that reward only the final output, our framework defines reward functions over the intermediate reasoning steps themselves, incentivizing chains that are not only correct but also coherent, structured, and interpretable [8], [9]. This dual optimization objective represents a significant departure from conventional LLM training paradigms and introduces new architectural and infrastructural considerations. The remainder of this paper explores these considerations in depth, beginning with a review of related work before detailing the proposed framework, examining its implementation challenges, and discussing its broader implications for AI governance and socio-technical system design.

2. Related Work and Theoretical Foundations

The challenge of LLM reasoning has been approached from multiple directions within the AI research community. Chain-of-thought prompting represents one of the most influential

contributions, demonstrating that asking models to generate intermediate reasoning steps before producing a final answer significantly improves performance on complex reasoning tasks [10]. This technique exploits the sequential nature of language generation to encourage models to externalize reasoning processes. Subsequent work has refined this approach through techniques such as self-consistency, where multiple reasoning paths are generated and aggregated, and tree-of-thought methods that explore alternative reasoning branches [11], [12]. While these methods enhance performance, they do not fundamentally alter the underlying reasoning architecture but rather modify the prompting strategy.

Reinforcement learning has been applied to LLMs primarily through techniques like reinforcement learning from human feedback, which aligns model outputs with human preferences through reward modeling [8]. This approach has proven remarkably effective for improving output quality and safety, but it typically operates at the level of entire responses rather than individual reasoning steps. The application of RL to intermediate reasoning processes remains comparatively underexplored, though promising directions include process reward models that evaluate each step in a reasoning chain [13]. Our framework extends this line of inquiry by explicitly incorporating dual-process theory into the reinforcement learning architecture.

Dual-process theories have a long history in cognitive science, with applications ranging from judgment and decision-making to social cognition and moral reasoning [7], [14]. The distinction between intuitive and analytical processing is not merely descriptive but has substantial implications for understanding cognitive biases, expertise development, and error correction. In the context of AI systems, several researchers have proposed hybrid architectures that combine neural networks with symbolic reasoning components to achieve more robust and interpretable behavior [4], [6]. These approaches, while intellectually compelling, have faced challenges in scaling to the complexity of modern language tasks.

The integration of dual-process theory with reinforcement learning for reasoning chain optimization is proposed in a recent study that explores high-level planning guidance for LLM reasoning [15]. This work demonstrates that providing hierarchical planning signals during reinforcement learning can improve both the correctness and the structural coherence of reasoning chains. Our framework builds upon this insight by formalizing the dual-process distinction and embedding it within a comprehensive cognitive architecture designed for explainability.

3. The Cognitive Chain Optimization Framework

The Cognitive Chain Optimization framework is built upon a fundamental architectural separation between two reasoning modalities operating within a unified language model. System One in this framework corresponds to a fast, associative processing pathway that generates candidate reasoning steps based on learned patterns and heuristics. This system is responsible for the initial generation of ideas, hypotheses, and potential inference steps without deep verification. In the context of LLMs, System One can be understood as the standard inference trajectory that the base model would follow when prompted directly, leveraging its vast associative memory to produce plausible continuations.

System Two, in contrast, is a deliberative processing pathway that engages in systematic verification, logical consistency checking, and error detection. This system operates more slowly and computationally expensively, but it provides the capacity for analytical oversight that System One lacks. Within the CCO framework, System Two does not generate reasoning

from scratch but rather evaluates, refines, and sometimes overrides the outputs of System One. The interaction between these two systems is governed by a meta-cognitive controller that determines when System Two should be engaged and how much computational resources should be allocated to verification.

The reinforcement learning component of the framework operates across this dual architecture. Rather than treating the entire reasoning chain as a monolithic action sequence, the RL agent learns to optimize the chain at multiple levels of granularity. At the highest level, the agent learns meta-cognitive policies that determine the allocation of computational resources between System One and System Two. At the intermediate level, the agent learns to generate planning structures that guide the overall direction of reasoning. At the lowest level, the agent learns to produce individual reasoning steps that are coherent, relevant, and aligned with the desired outcome [15], [16].

The reward structure for this RL system is multi-faceted. Correctness rewards are assigned for reasoning chains that lead to accurate final answers. Coherence rewards are assigned for chains that maintain logical consistency across steps. Explainability rewards, a novel contribution of this framework, are assigned for chains that are structurally transparent, meaning that each step can be clearly understood and justified by human evaluators. This triple reward structure creates optimization pressures that differ markedly from standard RL approaches, encouraging the model to develop reasoning strategies that are not merely effective but also inherently interpretable.

Training such a system presents significant computational challenges. The dual-process architecture introduces additional parameters and inference pathways, increasing the memory and compute requirements for both training and deployment. However, we argue that this investment is justified by the substantial gains in explainability and trustworthiness. Moreover, the meta-cognitive controller can be designed to minimize overhead in routine cases, engaging System Two only when necessary based on confidence estimates or complexity metrics. This adaptive resource allocation is itself a learned behavior within the RL framework, creating a system that becomes more efficient over time.

4. Structural Trade-offs and Architectural Considerations

The implementation of a dual-process reasoning architecture within LLMs introduces a series of structural trade-offs that must be carefully balanced. The most fundamental trade-off is between reasoning depth and computational efficiency. System Two processes, while essential for verification and error correction, consume significantly more computational resources than System One processes. A system that engages System Two excessively will be slow and expensive to operate, undermining the practical utility of the model. Conversely, a system that relies too heavily on System One will produce reasoning chains that are plausible but potentially flawed, compromising both accuracy and explainability.

Addressing this trade-off requires sophisticated meta-cognitive policies that can dynamically adjust the balance between the two systems based on task characteristics and context. The reinforcement learning framework is well-suited for learning such policies, as it can optimize for a combination of accuracy, speed, and resource consumption over the long term. However, the learning process itself introduces additional complexity, as the meta-cognitive controller must be trained alongside the reasoning systems themselves, creating a challenging multi-agent optimization problem.

Another critical trade-off concerns the degree of separation between the two reasoning pathways. A fully modular architecture, in which System One and System Two are implemented as entirely separate models or modules, offers clean interfaces and independent optimization opportunities but sacrifices the benefits of shared representations. A more integrated architecture, in which the two systems share most parameters and differ only in their processing modes, offers greater efficiency and transferability but makes it more difficult to understand and control their interactions [17]. The CCO framework adopts a hybrid approach, with shared embedding and attention layers but distinct processing pathways that can be selectively activated.

The architectural design also has implications for model governance and auditing. A modular architecture with clearly defined interfaces between System One and System Two allows external auditors to inspect the behavior of each system independently, potentially identifying biases or failure modes in the heuristic generation process that might be invisible in the final output. Similarly, the traceability of reasoning chains through the dual system enables fine-grained accountability, where specific errors can be attributed to particular processing steps rather than being buried in the model's overall behavior [3], [18]. This traceability is essential for deploying LLMs in regulated environments where decision-making processes must be documented and reviewable.

5. Explainability, Transparency, and Human Alignment

The primary motivation for the Cognitive Chain Optimization framework is the production of reasoning chains that are intrinsically explainable. Explainability in this context means more than simply providing a textual justification for a conclusion. It means that the reasoning process itself is structured in a way that is accessible to human understanding, that the logical connections between steps are clear, and that the model's confidence in its own reasoning can be assessed at each stage [5], [6]. The dual-process architecture contributes to this goal by making explicit the distinction between heuristic generation and analytical verification, allowing human observers to see where intuitive leaps occur and how they are subsequently evaluated.

The structure of reasoning chains produced by the CCO framework supports multiple levels of explanation. At a coarse level, the chain can be summarized as a sequence of high-level planning steps, showing the overall trajectory of the reasoning process. At a finer level, each individual step can be examined, revealing the specific inferences and verifications that were performed. This hierarchical transparency is valuable both for end users who need to understand the basis for a model's conclusion and for developers and auditors who need to diagnose and improve the model's behavior [19].

Alignment with human reasoning patterns is another important dimension of explainability. The dual-process framework is explicitly inspired by human cognition, which means that reasoning chains generated by the CCO model should be more intuitive for human observers to follow and evaluate. This alignment reduces the cognitive burden on users who must assess the trustworthiness of model outputs and facilitates more effective human-AI collaboration in decision-making contexts. However, it is important to acknowledge that human reasoning itself is subject to biases and limitations, and that aligning AI reasoning with human patterns may perpetuate those biases if not carefully managed [14], [20].

The explainability properties of the CCO framework also have implications for fairness and bias mitigation. By making the reasoning chain transparent, it becomes possible to identify

points at which biased heuristics might influence the reasoning process and to design interventions that target those specific stages. For example, if System One consistently generates stereotyping associations in certain contexts, System Two can be trained to detect and correct those associations before they propagate through the reasoning chain [21]. This approach to bias mitigation is more targeted and potentially more effective than post-hoc debiasing methods that modify the model's final outputs without addressing the underlying reasoning processes.

6. Infrastructure, Deployment, and Sustainability Considerations

Deploying a dual-process reasoning framework at scale presents significant infrastructural challenges that must be addressed for practical adoption. The most immediate challenge is computational cost, as System Two processing introduces additional inference steps that can multiply the required compute resources by a significant factor. This cost is not uniform across all queries but depends on the complexity of the task and the decisions of the meta-cognitive controller. From an infrastructure perspective, this variability requires flexible resource allocation mechanisms that can handle unpredictable spikes in compute demand without degrading overall system performance [22].

One approach to managing computational costs is hierarchical caching, where common reasoning patterns and intermediate results are cached for reuse across multiple queries. If System Two frequently performs similar verification checks on similar types of inputs, those verification results can be stored and retrieved rather than recomputed each time. Similarly, System One's heuristic generation can benefit from caching of common patterns. This caching strategy can significantly reduce the average computational cost per query while maintaining the benefits of dual-process reasoning for novel or complex cases [23].

Context-aware resource allocation is another important infrastructure consideration. The meta-cognitive controller can be designed to consider not only the complexity of the current task but also the available computational resources and response time requirements. In a deployment scenario where low latency is critical, the controller may choose to rely more heavily on System One processing, accepting some reduction in explainability for the sake of speed. In scenarios where accuracy and auditability are paramount, the controller can allocate additional resources to System Two verification. This flexibility allows the same model to serve diverse deployment contexts with appropriate trade-offs [24].

Sustainability is an increasingly important concern in AI system design, and the computational overhead of dual-process reasoning must be evaluated in this context. The additional compute required for System Two processing translates directly into increased energy consumption and carbon emissions. However, this cost must be weighed against the benefits of explainability, which can reduce the need for repeated queries, enable more efficient debugging, and facilitate regulatory compliance. A life-cycle assessment of the framework's environmental impact should consider not only the operational costs but also the potential for reduced model retraining and more targeted improvement efforts made possible by the transparency of the reasoning process [25].

7. Robustness, Fairness, and Governance Implications

The robustness of reasoning chains produced by the CCO framework is a critical consideration for deployment in high-stakes applications. The dual-process architecture provides inherent redundancy, as System Two can catch errors that System One might generate. However, this redundancy is only valuable if the two systems have complementary

failure modes. If both systems share similar vulnerabilities, such as susceptibility to the same types of adversarial inputs, then the dual-process architecture may provide limited robustness benefits. Designing the two systems to have genuinely different failure profiles requires careful architectural choices and training strategies [26].

Bias amplification is a well-documented risk in AI systems, and the introduction of dual-process reasoning does not automatically mitigate this risk. In fact, the additional complexity of the framework could introduce new sources of bias if not carefully managed. For example, the meta-cognitive controller might learn to allocate more System Two resources to inputs from certain demographic groups, creating disparities in reasoning quality across groups. Alternatively, System One's heuristic generation might systematically fail for underrepresented contexts, and System Two might not be trained to detect such failures. Addressing these risks requires explicit fairness objectives in the reinforcement learning reward structure and ongoing monitoring of model behavior across diverse input distributions [21], [27].

The governance implications of deployable explainable AI systems extend beyond technical design to encompass institutional frameworks, regulatory compliance, and stakeholder engagement. Reasoning chains produced by the CCO framework, by virtue of their transparency, can serve as auditable records of decision-making processes. This auditability is valuable for regulatory compliance in domains such as healthcare, finance, and criminal justice, where decisions must be justifiable and reviewable. However, auditability also raises questions about privacy and data protection, as the reasoning chains may contain information about the input data that could be sensitive or proprietary [28].

Standardization of explainability metrics and evaluation protocols is an important governance challenge that the CCO framework can help address. Because the framework produces structured reasoning chains with clear separation between heuristic and analytical processing, it enables more systematic evaluation of reasoning quality than is possible with black-box models. Standardized benchmarks for measuring the correctness, coherence, and explainability of reasoning chains can facilitate comparison across different approaches and support the development of best practices for deployment [29]. Such standardization is a prerequisite for the kind of regulatory oversight that is likely to emerge as AI systems become more deeply embedded in critical infrastructure.

8. Conclusion and Future Directions

Cognitive Chain Optimization represents a significant departure from conventional approaches to LLM reasoning, proposing a framework that is explicitly designed for explainability from the ground up rather than treating it as an afterthought. By integrating dual-process cognitive theory with reinforcement learning, CCO creates reasoning chains that are both structurally coherent and interpretable, offering substantial advantages for high-stakes applications where trust, accountability, and auditability are paramount. The framework addresses fundamental trade-offs between reasoning depth, computational efficiency, and transparency, providing architectural mechanisms for dynamically balancing these competing objectives.

The broader implications of this work extend beyond technical performance to encompass governance, sustainability, and socio-technical system design. As AI systems become increasingly integrated into institutional decision-making processes, the ability to produce auditable reasoning chains will become not merely a desirable feature but a regulatory

requirement. The CCO framework provides a foundation for meeting these requirements while maintaining the flexibility and efficiency that make LLMs valuable in practice.

Future research directions include the development of more sophisticated meta-cognitive controllers that can learn to allocate computational resources with greater efficiency and adaptivity. The interaction between System One and System Two in more open-ended reasoning tasks, such as creative problem-solving or strategic planning, presents interesting theoretical and empirical challenges. Extending the framework to multimodal reasoning, where inputs and outputs span text, images, and other modalities, will be essential for deploying CCO in real-world applications. Longitudinal studies of deployed CCO systems, examining their performance, robustness, and fairness over time, will provide valuable empirical evidence to guide further development. Finally, the governance and policy dimensions of explainable reasoning systems merit sustained interdisciplinary attention, as the technical capabilities of these systems will increasingly shape the regulatory landscape in which they operate.

References

1. Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

3. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (pp. 59–68).

4. Garcez, A. d., & Lamb, L. C. (2023). Neurosymbolic AI: The 3rd wave. Artificial Intelligence Review, 56(3), 2235–2264.

5. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.

6. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312.

7. Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annual Review of Psychology, 59, 255–278.

8. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv preprint arXiv:2203.02155.

9. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

10. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. arXiv preprint arXiv:2201.11903.

11. Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2023). Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171.

12. Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. arXiv preprint arXiv:2305.10601.

13. Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., & Cobbe, K. (2023). Let's verify step by step. arXiv preprint arXiv:2305.20050.

14. Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences, 23(5), 645–665.

15. Dou, Z., Zhao, Q., Wan, Z., Zhang, D., Wang, W., Raiyan, T., ... & Biswas, S. (2025). Plan Then Action: High-Level Planning Guidance Reinforcement Learning for LLM Reasoning. arXiv preprint arXiv:2510.01833.

16. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

17. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1–67.

18. Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62.

19. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

20. Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131.

21. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.

22. Barroso, L. A., & Hölzle, U. (2009). The datacenter as a computer: An introduction to the design of warehouse-scale machines. Synthesis Lectures on Computer Architecture, 4(1), 1–108.

23. Potter, S. S., & Koch, R. (2023). Caching strategies for large-scale machine learning inference. Proceedings of the 20th USENIX Symposium on Networked Systems Design and Implementation, 345–360.

24. Huang, Y., Chen, L., & Zhang, Z. (2022). Context-aware resource allocation for adaptive AI inference. IEEE Transactions on Parallel and Distributed Systems, 33(11), 2897–2910.

25. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650).

26. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

27. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 77–91).

28. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

29. Lipton, Z. C. (2018). The mythos of model interpretability. Queue, 16(3), 31–57.